PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :		(11) International Publication Number: WO 98/11236
C12N 15/57, 9/64, 5/10, 15/11, A61K 38/48, C12Q 1/37, C07K 16/40, A61K 39/395, 31/70	A1	(43) International Publication Date: 19 March 1998 (19.03.98)

(21) International Application Number: PCT/GB97/02426

(22) International Filing Date: 9 September 1997 (09.09.97)

(30) Priority Data:
9618966.7 11 September 1996 (11.09.96) GB

(71) Applicants *(for all designated States except US)*: SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB). UNIVERSITY OF WALES [GB/GB]; Cardiff CF1 3TE (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: POWELL, David [IE/US]; SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989 (US). KAY, John [GB/GB]; University of Wales, Cardiff CF1 3TE (GB). HILL, Jeffrey [GB/GB]; University of Wales, Cardiff CF1 3TE (GB).

(74) Agent: CONNELL, Anthony, Christopher; SmithKline Beecham, Corporate Intellectual Property, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ASPARTIC PROTEASE

(57) Abstract

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides. More particularly, the polypeptide of the present invention is believed to be a novel aspartic protease. The invention also relates to inhibiting the action of such polypeptides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

ASPARTIC PROTEASE

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides. More particularly, the polypeptide of the present invention is believed to be a novel aspartic protease. The invention also relates to inhibiting the action of such polypeptides.

There are currently five known human aspartic proteases, namely pepsin, gastricsin, cathespin D, cathespin E and renin, and these have widely varying functions. Pepsin and gastricsin are involved in nutritive processes in the stomach, cathepsin D is involved in protein turnover in nearly all cell types, and renin has the highly specific function of angiotensin production from its precursor form, angiotensinogen. The precise role of cathepsin E remains to be confirmed, although its location in some epithelial cells types has indicated a role in antigen processing. It may also be involved in certain inflammatory conditions, e.g. *Helicobacter pylori* infection in the stomach.

There is the possibility that a novel aspartic protease could have an extracellular function. There are at present a number of such processes which are thought to involve aspartic proteases, but where the exact enzyme involved remains to be identified. Important functions implicated are the processing of endothelin and pro-opiomelanocortin prohormones. An aspartic protease is also thought to be involved in the processing of the serum amyloid A protein.

Aspartic proteases are so called because of the pair of aspartic acid (Asp) residues that are required for the hydrolytic cleavage of a peptide bond. The catalytic Asp residues are normally located within a -Hyd-Hyd-Asp-Thr-Gly- active site motif sequence (where Hyd can be any hydrophobic amino acid). Eucaryotic aspartic proteases usually possess two such sequences, which have been shown by crystallography to lie adjacent to each other in the active site of the enzyme. Another highly conserved part of these enzymes comprises a structural beta-hairpin loop (often termed the "flap") which lies over the active site and may assist in substrate binding.

A novel aspartic protease was described by Jordan Tang *et al*, Okalahoma Medical Research Foundation during a presentation at the VII International Conference on Aspartic Proteinases Oct 22-27 1996.

In accordance with one aspect of the present invention, there is provided a novel polypeptide which comprises the amino acid sequence given in SEQ ID NO 1 or SEQ ID NO 15, or a fragment, analog or derivative thereof. The polypeptide of the present invention is of human origin. The polypeptide is believed to be an aspartic protease.

Polypeptides of the present invention further include a polypeptide having the amino acid sequence contained in SEQ ID NO 1 or 15; and a polypeptide comprising an amino acid

sequence which has at least 80%, preferably at least 90%, more preferably 95%, still more preferably at least 97 to 99% identity to any of the amino acid sequences SEQ ID NO 1 or 15 over its entire length. Also included are polypeptides having the amino acid sequence which have at least 80%, preferably at least 90%, more preferably 95%, still more preferably at least 97 to 99% identity to any of the amino acid sequences SEQ ID NO 1 or 15 over its entire length. Polypeptides of the present invention also include a polypeptide comprising an amino acid sequence encoded by a polynucleotide that have at least 80% identity to that of SEQ ID NO:2 or 16 over its entire length.

The polypeptides may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

Fragments are also included in the invention. A fragment is a polypeptide having an amino acid sequence that entirely is the same as part, but not all, of the amino acid sequence of the aforementioned polypeptides. Such fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, and 101 to the end of polypeptide of SEQ ID N0 1 or 15. In this context "about" includes the particularly recited ranges larger or smaller by several, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of SEQ ID N0 1 or 15, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Other preferred fragments are biologically active fragments. Biologically active fragments are those that mediate biological activity, for instance aspartyl protease activity, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

Preferably, all of these polypeptide fragments retain the biological activity of the precursor polypeptide, including antigenic activity. Variants of the defined sequence and fragments also form part of the present invention. Preferred variants are those that vary from

the referents by conservative amino acid substitutions -- i.e., those that substitute a residue with another of like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In accordance with another aspect of the present invention, there are provided polynucleotides (DNA or RNA) which encode such polypeptides.

In accordance with a preferred aspect of the present invention, there is provided a polynucleotide which encodes for the polypeptide having the amino acid sequence of SEQ ID NO 1 or SEQ ID NO 15.

In particular, the invention provides a polynucleotide having the DNA sequence given in SEQ ID NO 2 or SEQ ID NO 16. The invention further provides a polynucleotide encoding a polypeptide which comprises the DNA sequence given in SEQ ID NO 2 or SEQ ID NO 16.

Polynucleotides of the present invention further include a polynucleotide comprising a nucleotide sequence that has at least 80%, preferably at least 90%, more preferably at least 95% identity over its entire length to a nucleotide sequence encoding a polypeptide of SEQ ID NO 1 or SED ID NO 15, and a polynucleotide comprising a nucleotide sequence that is at least 80%, preferably at least 90%, more preferably at least 95% identical to that of SEQ ID NO:2 or SEQ ID NO 16 over its entire length. Furthermore, those with at least 97% are highly preferred and those with at least 98-99% are most highly preferred, with at least 99% being the most preferred.

cDNA molecules (ESTs) showing extended identity sections with the cDNA of SEQ ID NO 2 and SEQ ID NO 16 have been identified in cDNA libraries of human origin from a wide variety of sources. These ESTs are given in SEQ ID NO 3 to 14:

Sequence No:	Library:
SEQ ID NO 3 (EST 176432):	Raji cells, cyclohexamide treated
SEQ ID NO 4 (EST 424772):	Human b cell lymphoma
SEQ ID NO 5 (EST 443275):	Breast lymph node cDNA library
SEQ ID NO 6 (EST 567394):	Raji cells, cyclohexamide treated
SEQ ID NO 7 (EST 685578):	Human activated monocytes
SEQ ID NO 8 (EST 928138):	Fetal Liver, subtraction II
SEQ ID NO 9 (EST 947785):	Breast lymph node cDNA library
SEQ ID NO 10 (EST 1000163):	Spinal chord

SEQ ID NO 11 (EST 1218021): Spleen, chronic lymphocytic leukemia
SEQ ID NO 12 (EST 1320439): Human tonsils, lib 2
SEQ ID NO 13 (EST 716478): cd34 depleted buffy coat cord blood
SEQ ID NO 14 (EST 857644): Human adult testis, large inserts

Accordingly, in a further aspect, the present invention provides for a polynucleotide which encodes for an aspartic protease characterised by one or more partial DNA sequences selected from SEQ ID NOs 3 to 14.

The polynucleotides of SEQ ID NO 2 and SEQ ID NO 16 are structurally related to the aspartic protease family. SEQ ID NO 2 and 16 have 1376 and 1347 nucleotides, respectively. The first (or N-Terminal) active site motif is believed to be -Ala-Phe-Asp-Thr-Gly- and is encoded by at least two ESTs (SEQ ID NO 11 and 12), by the identical DNA sequence in each case (-GCC-TTT-GAC-ACT-GGC-). The second (or C-Terminal) active site motif is believed to be -Ile-Leu-Asp-Thr-Gly and is encoded (by the DNA sequence -ATC-CTG-GAT-ACA-GGC-) by at least one EST (SEQ ID NO 13). The conserved flap region of the enzyme is believed to be -Tyr-Gly-Thr-Gly-.and is encoded by at least four ESTs (SEQ ID NO 9, 10, 5 and 12) with the same DNA sequence (-TAT-GGA-ACT-GGG-) in all cases. The sequences SEQ ID NO 1 and SEQ ID NO 15 are believed to contain the entire length of the mature form of the novel aspartic protease, with a further approximately 60 amino acids comprising the propart (or prosegment). The initiating Met residue may also be identified in SEQ ID NO 1, the ORF starting at nucleotide 26 . Simple blast sequence analysis reveals that the polypeptide of SEQ ID NO 15 is most homologous to Prepro-Cathepsin D from chicken. It is however believed that polypeptides of SEQ ID NO 1 and SEQ ID NO 15 will be, at a functional level, most like Cathepsin E due to the presence of a Cathepsin E specific glycosylation site (-Asn-Phe-Thr-) just before the sequence encoding the N-Terminal active site motif. The widespread distribution in many cell types of ESTs associated with the polypeptide may indicate a non-specific hydrolytic function (similar to cathepsin E and cathepsin D) for the encoded poypetide.

The polynucleotide of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the polypeptide may be identical to the coding sequence shown in SEQ ID NO 2 or SED ID NO 16 or may be a different coding sequence which, as a result of the redundancy or degeneracy of the genetic code, encodes the same polypeptide as the DNA of SEQ ID NO 2 or SED ID NO 16.

The present invention includes variants of the hereinabove described polynucleotides which encode fragments, analogs and derivatives of the polypeptide having the amino acid sequence of SEQ ID NO 1 or SED ID NO 15. The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same polypeptide as shown in SEQ ID NO 1 or SED ID NO 15, as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the polypeptide. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

The polynucleotide may have a coding sequence which is a naturally occurring allelic variant of the coding sequence of SEQ ID NO 2 or SED ID NO 16. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The polynucleotide which encodes for the polypeptide of SEQ ID NO 1 or SED ID NO 15 may include: only the coding sequence for the polypeptide; the coding sequence for the polypeptide and additional coding sequence such as a leader or secretory sequence or a proprotein sequence; the coding sequence for the polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature polypeptide.

Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention therefore includes polynucleotides, wherein the coding sequence for the polypeptide may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell. The polypeptide having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the polypeptide. The polynucleotides may also encode for a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a proprotein and is an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode for a mature protein, or for a protein having a prosequence or for a protein having both a prosequence and a presequence (leader sequence).

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexa-histidine tag supplied by a pQE-9 vector to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 50% and preferably 70% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides . As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode polypeptides which retain substantially the same biological function or activity as the polypeptide of SEQ ID NO 1 or SED ID NO 15.

The terms "fragment," "derivative" and "analog" when referring to the polypeptide of SEQ ID NO 1 or SED ID NO 15, means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

As used herein, the term "identity" is a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, 1993; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, 1987; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCS program package (Devereux, J., *et al.*, *Nucleic Acids Research* (1984) 12(1):387), BLASTP, BLASTN, FASTA (Atschul, S.F. *et al.*, *J Molec Biol* (1990) 215:403).

As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence of SEQ ID NO: 2/16 is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per

each 100 nucleotides of the reference nucleotide sequence of SEQ ID NO: 2/16. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5 or 3 terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference amino acid sequence SEQ ID NO 1/15 is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide, preferably a recombinant polypeptide.

The fragment, derivative or analog of the polypeptide of SEQ ID NO 1 or or SED ID NO 15 may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring

polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment. The polypeptide is preferably in purified form. By purified form is meant at least 80%, more preferably 90%, still more preferably 95% and most preferably 99% pure with respect to other protein contaminants.

The DNA of the present invention also makes possible the development by homologous recombination or "knockout" stategies (Kapecchi, Science, 244,:1288-1292 (1989) of animals that fail to express, or express a variant form of this enzyme

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

In accordance with yet a further aspect of the present invention, there is therefore provided a process for producing the polypeptide of the invention by recombinant techniques by expresssing a polynucleotide encoding said polypeptide in a host and recovering the expressed product. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a cosmid, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

Suitable expression vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli. lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression

of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The gene can be placed under the control of a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator (collectively referred to herein as "control" elements), so that the DNA sequence encoding the desired protein is transcribed into RNA in the host cell transformed by a vector containing this expression construction. The coding sequence may or may not contain a signal peptide or leader sequence. The protein sequences of the present invention can be expressed using, for example, the *E. coli* tac promoter or the protein A gene (spa) promoter and signal sequence. Leader sequences can be removed by the bacterial host in post-translational processing. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are PKK232-8 and PCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In addition to control sequences, it may be desirable to add regulatory sequences which allow for regulation of the expression of the protein sequences relative to the growth of the host cell. Regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may also be present in the vector, for example, enhancer sequences.

An expression vector is constructed so that the particular coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the control sequences being such that the coding sequence is transcribed under the "control" of the control sequences (i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence). Modification of the coding sequences may be desirable to achieve this end. For example, in some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame. The control sequences and other regulatory sequences may be ligated to the coding sequence prior to insertion into a vector, such as the cloning vectors described above. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site. Modification of the coding

sequences may also be performed to alter codon usage to suit the chosen host cell, for enhanced expression.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

Examples of recombinant DNA vectors for cloning and host cells which they can transform include the bacteriophage λ (*E. coli*), pBR322 (*E. coli*), pACYC177 (*E. coli*), pKT230 (gram-negative bacteria), pGV1106 (gram-negative bacteria), pLAFR1 (gram-negative bacteria), pME290 (non-*E. coli* gram-negative bacteria), pHV14 (*E. coli* and *Bacillus subtilis*), pBD9 (*Bacillus*), pIJ61 (*Streptomyces*), pUC6 (*Streptomyces*), YIp5 (*Saccharomyces*), a baculovirus insect cell system, , YCp19 (*Saccharomyces*). See, generally, "DNA Cloning": Vols. I & II, Glover *et al.* ed. IRL Press Oxford (1985) (1987) and; T. Maniatis *et al.* ("Molecular Cloning" Cold Spring Harbor Laboratory (1982).

In some cases, it may be desirable to add sequences which cause the secretion of the polypeptide from the host organism, with subsequent cleavage of the secretory signal.

Yeast expression vectors are also known in the art. See, e.g., U.S. Patent Nos. 4,446,235; 4,443,539; 4,430,428; see also European Patent Applications 103,409; 100,561; 96,491. pSV2neo (as described in J. Mol. Appl. Genet. 1:327-341) which uses the SV40 late promoter to drive expression in mammalian cells or pCDNA1neo, a vector derived from pCDNA1(Mol. Cell Biol. 7:4125-29) which uses the CMV promoter to drive expression. Both these latter two vectors can be employed for transient or stable(using G418 resistance) expression in mammalian cells. Insect cell expression systems, e.g., *Drosophila*, are also useful, see for example, PCT applications WO 90/06358 and WO 92/06212 as well as EP 290,261-B1.

Polypeptides can be expressed in host cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by

Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples including the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

In a further aspect, the present invention relates to host cells containing the above-described vectors. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. As representative examples of appropriate hosts, there may be mentioned: prokaryotes for example bacterial cells, such as *E. coli*, *Streptomyces*, *Salmonella typhimurium* and eukaryotes for example fungal cells, such as yeast, insect cells such as *Drosophila* and *Spodoptera frugiperda*, mammalian cells such as CHO, COS or Bowes melanoma, plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation. (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Depending on the expression system and host selected, the polypeptide of the present invention may be produced by growing host cells transformed by an expression vector described above under conditions whereby the polypeptide of interest is expressed. The polypeptide is then isolated from the host cells and purified. If the expression system secretes the polypeptide into growth media, the polypeptide can be purified directly from the media. If the polypeptide is not secreted, it is isolated from cell lysates or recovered from the cell membrane fraction. Where the polypeptide is localized to the cell surface, whole cells or isolated membranes can be used as an assayable source of the desired gene product. Polypeptide expressed in bacterial hosts such as *E. coli* may require isolation from inclusion bodies and refolding. The selection of the appropriate growth conditions and recovery methods are within the skill of the art.

The polypeptide can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue.

The polypeptide of the present invention is also useful for identifying other molecules which have similar biological activity. An example of a screen for this is isolating the coding region of the aspartic protease gene by using the known DNA sequence to synthesize an oligonucleotide probe or as a probe itself. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of human cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polypeptides may also be employed in accordance with the present invention by expression of such polypeptides *in vivo*, which is often referred to as "gene therapy."

Thus, for example, cells from a patient may be engineered with a polynucleotide (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being provided to a patient to be treated with the polypeptide. Such methods are well-known in the art. For example, cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by, for example, procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding the polypeptide of the present invention may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in*

vivo. These and other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells in vivo after combination with a suitable delivery vehicle.

"Recombinant" polypeptides refer to polypeptides produced by recombinant DNA techniques; i.e., produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide. "Synthetic" polypeptides are those prepared by chemical synthesis.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control.

A "vector" is a replicon, such as a plasmid, phage, or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "double-stranded DNA molecule" refers to the polymeric form of deoxyribonucleotides (bases adenine, guanine, thymine, or cytosine) in a double-stranded helix, both relaxed and supercoiled. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the sense strand of DNA.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular protein, is a DNA sequence which is transcribed and translated into a polypeptide when placed under the control of appropriate regulatory sequences.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

DNA "control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the expression (i.e., the transcription and translation) of a coding sequence in a host cell.

A control sequence "directs the expression" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

A "host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous DNA sequence.

A cell has been "transformed" by exogenous DNA when such exogenous DNA has been introduced inside the cell membrane. Exogenous DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes and yeasts, for example, the exogenous DNA may be maintained on an episomal element, such as a plasmid. With respect to eukaryotic cells, a stably transformed or transfected cell is one in which the exogenous DNA has become integrated into the chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cell containing the exogenous DNA.

A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

Two DNA or polypeptide sequences are "substantially homologous" or "substantially the same" when at least about 85% (preferably at least about 90%, and most preferably at least about 95%) of the nucleotides or amino acids match over a defined length of the molecule and includes allelic variations. As used herein, substantially homologous also refers to sequences showing identity to the specified DNA or polypeptide sequence. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., "Current Protocols in Mol. Biol." Vol. I & II, Wiley Interscience. Ausbel *et al.* (ed.) (1992). Protein sequences that are substantially the same can be identified by proteolytic digestion, gel electrophoresis and microsequencing.

The term "functionally equivalent" intends that the amino acid sequence of the subject protein is one that will exhibit enzymatic activity of the same kind as that of the aspartic protease.

A "heterologous" region of a DNA construct is an identifiable segment of DNA within or attached to another DNA molecule that is not found in association with the other molecule in nature.

The polypeptides of the present invention may be be of use in therapy. Accordingly, in a further aspect, the present invention provides a polypeptide having the amino acid sequence given in SEQ ID NO 1 or SED ID NO 15, and fragments, analogs or derivative thereof, for use in therapy. Suitably, such polypeptides may play a role in preventing, ameliorating or correcting dysfunctions or diseases, including, but not limited to, hypertension, inflammation, asthma and cardio-pulmonary conditions.

The polypeptides and polynucleotides of the present invention may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a

therapeutically effective amount of the active agent, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

The pharmaceutical compositions may be administered in a convenient manner such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The polypeptides or polynucleotides of the present invention is administered in an amount which is effective for treatment and/or prophylaxis of the specific indication. The amounts and dosage regimens of active agent administered to a subject will depend on a number of factors such as the mode of administration, the nature of the condition being treated and the judgment of the prescribing physician.

The sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the cDNA is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clones to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. FISH requires use of the clones from which the EST was derived, and the longer the better. For example, 2,000 bp is good, 4,000 is better, and more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this technique, see Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York (1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb).

Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that cDNA sequence. Ultimately, complete sequencing of genes from several individuals is required to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

The polypeptides of the invention or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such

antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention.

This invention further provides a method of screening compounds to identify those compounds which inhibit the polypeptide comprising contacting isolated polypeptide with a test compound and measuring the rate of turnover of an enzyme substrate as compared with the rate of turnover in the absence of test compound. The invention also relates to compounds identified thereby.

This invention also provides transgenic non-human animals comprising a polynucleotide encoding a polypeptide of the invention. Also provided are methods for use of said transgenic animals as models for mutation and SAR (structure/activity relationship) evaluation as well as in drug screens.

The present invention is also directed to inhibitor molecules of the polypeptides of the present invention, and their use in reducing or eliminating the function of the polypeptide.

An example of an inhibitor is an antibody or in some cases, an oligonucleotide which binds to the polypeptide.

An example of an inhibitor is an antisense construct prepared using antisense technology. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. For example, the 5' coding portion of the polynucleotide sequence, which encodes for the polypeptides of the present invention, is used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (triple helix -see Lee et al., Nucl. Acids Res., 6:3073 (1979); Cooney et al, Science, 241:456 (1988); and Dervan et al., Science, 251: 1360 (1991)), thereby preventing transcription and the production of polypeptide. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of the mRNA molecule into the polypeptide (Okano, J. Neurochem., 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression. CRC Press, Boca Raton, FL (1988)). The oligonucleotides

described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of polypeptide.

Another example of an inhibitor is a small molecule which binds to and occupies the catalytic site of the polypeptide thereby making the catalytic site inaccessible to substrate such that normal biological activity is prevented. Examples of small molecules include but are not limited to small peptides or peptide-like molecules.

When used in therapy, the inhibitors of the invention are formulated in accordance with standard pharmaceutical practice.

The inhibitors which are active when given orally can be formulated as liquids, for example syrups, suspensions or emulsions, tablets, capsules and, lozenges.

A liquid formulation will generally consist of a suspension or solution of the compound or pharmaceutically acceptable salt in a suitable liquid carrier(s) for example, ethanol, glycerine, non-aqueous solvent, for example polyethylene glycol, oils, or water with a suspending agent, preservative, flavouring or colouring agent.

A composition in the form of a tablet can be prepared using any suitable pharmaceutical carrier(s) routinely used for preparing solid formulations. Examples of such carriers include magnesium stearate, starch, lactose, sucrose and cellulose.

A composition in the form of a capsule can be prepared using routine encapsulation procedures. For example, pellets containing the active ingredient can be prepared using standard carriers and then filled into a hard gelatin capsule; alternatively, a dispersion or suspension can be prepared using any suitable pharmaceutical carrier(s), for example aqueous gums, celluloses, silicates or oils and the dispersion or suspension then filled into a soft gelatin capsule.

Typical parenteral compositions consist of a solution or suspension of the compound or pharmaceutically acceptable salt in a sterile aqueous carrier or parenterally acceptable oil, for example polyethylene glycol, polyvinyl pyrrolidone, lecithin, arachis oil or sesame oil. Alternatively, the solution can be lyophilised and then reconstituted with a suitable solvent just prior to administration.

A typical suppository formulation comprises a compound of formula (I) or a pharmaceutically acceptable salt thereof which is active when administered in this way, with a binding and/or lubricating agent such as polymeric glycols, gelatins or cocoa butter or other low melting vegetable or synthetic waxes or fats.

Preferably the composition is in unit dose form such as a tablet or capsule.

Each dosage unit for oral administration contains preferably from 1 to 250 mg (and for parenteral administration contains preferably from 0.1 to 25 mg) of an inhibitor of the invention.

The daily dosage regimen for an adult patient may be, for example, an oral dose of between 1 mg and 500 mg, preferably between 1 mg and 250 mg, or an intravenous, subcutaneous, or intramuscular dose of between 0.1 mg and 100 mg, preferably between 0.1

mg and 25 mg, of the compound of the formula (I) or a pharmaceutically acceptable salt thereof calculated as the free base, the compound being administered 1 to 4 times per day. Suitably the compounds will be administered for a period of continuous therapy.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units to T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

Sequence Data

SEQ NO ID 1

MSPPPLLLPLLLLLPLLNVEPAGATLIRIPLRQVHPGRRTLNLLRGWGKP
AELPKLGAPSPGDKPASVPLSKFLDAQYFGEIGLGTPPQNFTVAFDTGSS
NLWVPSRRCHFFSVPCWFHHRFNPNASSSFKPSGTKFAIQYGTGRVDGIL
SEDKLTIGGIKGASVIFGEALWESSLVFTVSRPDGILGLGFPILSVEGVR
PPLDVLVEQGLLDKPVFSFYFNRDPEVADGGELVLGGSDPAHYIPPLTFV
PVTVPAYWQIHMERVKVGSRLTLCAQGCAAILDTGTPVIVGPTEEIRALH
AAIGGIPLLAGEYIIRCSEIPKLPAVSLLIGGVWFNLTAQDYVIQFAQGD
VRLCLSGFRALDIASPPVPVWILGDVFLGAYVTVFDRGDMKSGARVGLAR
ARPRGADLGRRETAQAQYRGCRPGDAHAHRVARASATQ

(438aa)

SEQ ID NO 2

TGGGTTCACACCCGGCTCCCCAGCGATGTCTCCACCACCGCTGCTGCTACCCTTGCTGCTGCTGCTGCC
TCTGCTGAATGTGGAGCCTGCTGGGGCCACACTGATCCGGATCCCTCTTCGTCAAGTCCACCCTGGACG
CAGGACCCTGAACCTACTGAGGGGATGGGGAAAACCAGCAGAGCTCCCCAAGTTGGGGGCCCCATCCCC
TGGGGACAAGCCTGCCTCGGTACCTCTCTCCAAATTCCTGGATGCCCAGTATTTTGGGGAAATTGGGCT
GGGAACGCCTCCACAAAACTTCACTGTTGCCTTTGACACTGGCTCCTCCAATCTCTGGGTCCCGTCCAG
GAGATGCCACTTCTTCAGTGTGCCCTGCTGGTTCCACCACCGCTTCAATCCCAATGCCTCCAGCTCCTT
CAAGCCCAGTGGGACCAAGTTTGCCATTCAGTATGGAACTGGGCGGGTAGATGGAATCCTGAGTGAGGA
CAAGCTGACTATTGGTGGAATCAAGGGTGCATCCGTGATTTTCGGGGAAGCTCTGTGGGAATCCAGCCT
GGTCTTCACTGTTTCCCGCCCCGATGGGATATTGGGCCTCGGTTTTCCCATTCTGTCTGTGGAAGGAGT
TCGGCCCCCGCTGGATGTACTGGTGGAGCAGGGGCTATTGGATAAGCCTGTCTTCTCCTTTTACTTCAA
CAGGGACCCTGAAGTGGCTGATGGAGGAGAGCTGGTCCTGGGGGGCTCAGACCCGGCACACTACATCCC
ACCCCTCACCTTCGTGCCAGTCACAGTCCCCGCCTACTGGCAGATCCACATGGAGCGTGTGAAGGTGGG
CTCACGGCTGACTCTCTGTGCCCAGGGCTGTGCTGCCATCCTGGATACAGGCACACCTGTCATCGTAGG
ACCCACTGAGGAGATCCGGGCCCTGCATGCAGCCATTGGGGGAATCCCCTTGCTGGCTGGGGAGTACAT
CATCCGGTGCTCAGAAATCCCAAAGCTCCCCGCAGTCTCACTCCTCATTGGGGGGGTCTGGTTTAATCT
CACGGCCCAGGATTACGTCATCCAGTTTGCTCAGGGTGACGTCCGCCTCTGCTTGTCCGGCTTCCGGGC
CTTGGACATCGCTTCGCCTCCAGTACCTGTGTGGATCCTCGGCGACGTTTTCTTGGGGGCGTATGTGAC
CGTCTTCGACCGCGGGGACATGAAGAGCGGCGCACGAGTGGGACTGGCGCGCGCTCGCCCTCGCGGAGC
GGACCTGGGAAGGCGCGAGACCGCGCAGGCGCAGTACCGCGGGTGCCGCCCAGGTGATGCGCATGCGCA
CCGGGTAGCc/aGAGCTAGCGCTACTCAGTAAAAATCCAATATTTCCATTGAAAAAAAAAAAAAAA

c/a=either a C or an A at this position, unknown at present
initiating ATG underlined, stop codon unknown

SEQ ID NO 3 (EST 176432)

NAATTCGGCANAGAAGGAAAACTAGGAAGCCTGGGTTCACACCCGGCTCCCCAGCAATGT
CTCCACCACTGCTGCTGCTACCCTTGACTGCTGCTGCTGCCTCTGCTGAATGTGGAGCCT
GCTGGGGCCACACTGATCCGGATCCCTCTTCGTCAAGTCCACCCTGGACGCAGGCCCCTG
AAACCTACTGAGGGGATGGGGAAAACCAGCAGAGCTCCCCAAGTTGGGGGGCCCCATCCC
CTGGGGACAAGCCTGCCTCGGTACCTCTNTCCAAATTCCTGGGTGGCCCATTATTTTGGG

GGAAATTGGGGTGGGGAACGGCTTCCACAAAATTTCATTGTTGGCTTTGNACATGGGTTN
CTTCCAATNTNTGGGTCCCGTTCCAGGAGANGGCCATTTTTTTNATGTTGGCCCTGNTGG
GTTCCACCACNGTTTAANTTCCAATGGNNTCCAATTTCTTTAAAGCCCATGGGGNCCAAN
TTTNGCCATTNAATNTGGGA

SEQ ID NO 4 (EST 424772):

GGAATCCTGAGTGAGGACAAGCTGACTGTTGGTGGANTCAAGGGTNCATCCGTGAATTTT
CGGGGAAGCTCTGTGGGAATCCAGCCTGGTCTTCACTGTTTCCCGNCCCGATGGGNATAT
TGGGCCTCGGTTTTCCCATTCTGTCTGTGGAAGGAGTTCGGCCCCCGCTGGGATGTACTG
GTGGAGCAGGGGCTATTGGATAAGCCTGTCTTCTCCTTTTACTTCAACAGGGACCCTGNA
AGTGGCTGATGGAGGAGAGCTGGTCCTGGGGGGCTCAGACCCGGCACACTACATCCCACC
CCTCACCTTCGTGNCCAGTTCACAGTTCCCCGGCTACTGGCAGATTCCACATGGTGCGTG
TTGAAGGTNGGGGTCAAGGGTGAANTNTNTGTGGCCCAGGGTGTGTTGGCCATCCTGGAT
AACAGCACAATTTTCATNGTAGGGACCCATTG

SEQ ID NO 5 (EST 443275):

TGGTTCCACCACCGCTTCAATCCCAATGCCTCCAGCTCCTTCAAGCCCAGTGGGACCAAG
TTTGCNATTCAGTATGGAACTGGGCGGGTAGATGGAATCCTGAAGTAAGGACAAGCTGAC
TATTGGTGGGAATCAAGGGTGCATCCGTGAATTTTGGGGGAAGCTCTGTGGGAATCCAGC
CTGGTCTTCACTGTTTCCCGCCCCGATGGGATATTGGGCCTCGGTTTTCCCATTCTGTCT
GTGGGAAGGAGTTCGGCCCCCGNTGGATGTTACTGGTGGGAGCANGGGGCTATTGGGTNA
AGCCCGTCTTCTNCCTTTTAANTTCAACAGGGGACCCTGAAAGTGGGTT

SEQ ID NO 6 (EST 567394):

NACCACTGCTGCTGCNACCCTTCCTGCTGCTGCTGCCTCTCCTGAATGTGGAGCCTGCTG
GGGCCACACTNATCCGNATCCCTCTTCGTNAAGTCCACCCTGGACGCAGGACCCTNAACC
TACTGAGGGGATGGGGAAAACCAGCAGAGCTCCCCAAGTTGGGGGCCCCATCCCCTGGGG
ACAAGCCTGCNTCGGTACCTCTNTCCAAATTCCTGGATGCCCAGTATTTTGGGGAAATTG
GGCTGGGAACGCCTCCACAAAACTTCACTGTTGCCTTTGAANACTGGCTCCTCCAATCTT
TGGGTCCCGTCCAGGTGTTGCCACTTGTTNCAGTGTGGCCCTGATTGGTTTNCACCCACC
NTTTTCAATTCCCATGNCCTTNCAG

SEQ ID NO 7 (EST 685578):

TCCACCACTGCTGCTGCTACCCTTNCTGCTGCTGCTGCCTCTNCTGAATGTGGAGCCTGC
TGGGGCCACACTGATCCGGNNCCCTCTTCGTNAAGTCCACCCTGGACGCAGGACCCTGAA
CCTACTGAGGGGATTGGGNAAANCAGCAGAGCTCGCCAAGTTGGGGGTCCNATNCCCTNG
GGACAAGGCTGGC

SEQ ID NO 8 (EST 928138):

GGGCAGNGGNTCTCCACCACTGCTGCTGCTACCCTTGCTGCTGCTGCTGCCTCTGCTGAA
TGTGGAGCCTGCTGGGGCCACACTGATCCGGATCCCTCTTCGTCAAGTCCACCCTGGACG
CAGGACCCTGAACCTACTGAGGGGATGGGGAAAACCAGCAGAGCTCCCCAAGTTNGGGGC
CCCATCCCCTNGGGACAAGCCTGCCTCGGTACCTCTCTTCAAATTCCTGGATGCCCAGTA
TTTTTGGGAAATTTGGCTTGGAACGCTTCACAAAACTTCACTGTTGCTTTGACAAT

SEQ ID NO 9 (EST 947785):

GGCACGAGCTNTTTCCATCACCACTTCAATCCCAATNCCTCCAGCTCCTTCAAGCCCAGT
GGGACCAAGTTTGCCATTCAGTATGGAACTGGGCGGGTAGATGGAATCCTGAGTGAGGAC
AAGCTGACTATTGGTGGAATCAAGGGTGCATCCGTGATTTTCGGGGAAGCTCTGTGGGAA
TCCAGCCTGGTCTTCACTGTTTCCCGNCCCGATGGGATATTGGGCCTCGGTTTTCCCATT
CTGTCTGTGGAAGGAGTTCGGCCCCCGCTGGATGTACTGGTGGAGCAAGGGCTATTGGAT
AAGCCTGTCTTCTTCCTTTAATTCAACAGGGNCCNGAAGTGGTTGANGGAGGAGAGCTG
GTCCCGGGGGGGTTCAAGACCCGGGAAATNACANCCCACCCT

SEQ ID NO 10 (EST 1000163):

GGCACGAGGAGATGCCACTTCTTCAGTGTGCCCTGCTGGTTCCACCACCGCTTCAATCCC
AATGCCTCCAGCTCCTTCAAGCCCAGTGGGACCAAGTTTGCCATTCAGTATGGAACTGGG
CGGGTAGATGGAATCCTGAGTGAGGACAAGCTGACTATTGGTGGAATCAAGGGTGCATCC
GTGATTTTCGGGGAAGCTCTGTGGGAATCCAGCCTGGTCTTCACTGTTTCCCGGCCCGAT
GGGATATTGGGNCTCGGTTTTCCCATTCTGTCTGTGGAAGGAGTTCGGNCCCCGCTGGAT
GTACTGGTGGAGCAGGGGNTATTGGATAAGCCTGTNTTCTTCTTTTAATTCAACAAGGAC
CCTGAAGTGGNTTAATGGAGGAGAGCTTGTCCTNGGGGGGNT

SEQ ID NO 11 (EST 1218021):

ACCCACGCGTCCGCACCACTGCTGCTGCTACCCTTGCTGCTGCTGCTGCCTCTGCTGAAT
GTGGAGCCTGCTGGGGCCACACTGATCCGGATCCCTCTTCGTCAAGTCCACCCTGGACGC
AGGACCCTGAACCTACTGAGGGGATGGGGAAAACCAGCAGAGCTCCCCAAGTTGGGGGCC
CCATCCCCTGGGGACAAGCCTGCCTCGGTACCTCTCTCCAAATTCCTGGATGCCCAGTAT
TTTGGGGAAATTGGGCTGGGAACGCCTCCACAAAACTTCACTGTTGCCTTTGACACTGGC
TCCTCCAATCTCTGGGTCCCGTCCAGGAGATGCCACTTCTTCAGTGTGCCCTGCTGGTTC
CAACAACGCTTCAATCCCAATGCCTCCAGCTCCTTCAAGCCCAGTGGGAACCAAGTNTGC
CATTCAGTATGGAACTNGGCCGGGTAGATGGGAATCCTGAATGANGACNAAGCT

SEQ ID NO 12 (EST 1320439):

GGGTCGACCCACGCGTCCGGGGCTGGGAACGCCTCCACAAAACTTCACTGTTGCCTTTGA
CACTGGCTCCTCCAATCTCTGGGTCCCGTCCAGGAGATGCCACTTCTTCAGTGTGCCCTG
CTGGTTCCACCACCGCTTCAATCCCAATGCCTCCAGCTCCTTCAAGCCCAGTGGGACCAA
GTTTGCCATTCAGTATGGAACTGGGCGGGTAGATGGAATCCTGAGTGAGGACAAGCTGAC
TATTGGTGGAATCAAGGGTGCATCCGTGATTTTCGGGGAAGCTCTGTGGGAATCCAGCCT
GGTCTTCACTGTTTCCCGCCCCGATGGGATATTGGGCCTCGGTTTTCCCATTCTGTCTGT
GGAAGGAGTTCGGCCCCCGCTGGATGTACTGGTGGAGCAAGGGCTATTGGATAAGCCTGT
CTTCTCCTTTTACTTCAACAGGGACCCTGAAAGTGGCTGATTGAAGAGAACTTGTCCTGG
G

SEQ ID NO 13 (EST 716478):

ACAGTCCCCNCCTACTGGCAGATCCACATGGAGCGTGTGAAGGTGGGCTCACGGCTGACT
CTCTGTGCCCAGGGCTGTGCTGCCATCCTGGATACAGGCACACCTGTCATCGTAGGACCC
ACTGAGGAGATCCGGGCCCTGCATGCAGCCATTGGGGGAATCCCCTTGCTGGCTGGGGAG
TACATCATCCGGTGCTCAGAAATCCCAAAGCTCCCCGCAGTCTCACTCCTCATTGGGGGG
GTCTGGTTTAATCTCACGGCCCAGGATTACGTCATCCAGTTTGCTCAGGGTGACGTCCGC
CTCTGCTTGTCCGGCTTCCGGGCCTTGGACATCGCTTGGCTNCAGTACCTGTGTGGATCC

TCGGCGACTTTTTNTTTCRRRRCGATGTGACCGTTTTGACCGGGGGACATNAAGAGCGGCG
AACGAGTNGACTTTCRRRRKKJGTTGCCTTGCGAGGACCTTGNAGCGGAACGGAAGGAGNCCG
GGGTCCCCAGTATCGATGACGGNGAAGTGGTCTAGAAACATATATTAAAAAAAAAAAAAA
AACTGGGNTTNGCCG

SEQ ID NO 14 (EST 857644):

GGCACGAGGCAAGGGTGCATCAGTGATTTTCGGGGAGGCTCTCTGGGAGCCCAGCCTGGT
CTTCGCTTTTGCCCATTTTGATGGGATATTGGGCCTCGGTTTTCCCATTCTGTCTGTGGA
AGGAGTTCGGCCCCCGATGGATGTACTGGTGGAGCAGGGGCTATTGGATAAGCCTGTCTT
CTCCTNTTACCTCAACAGGGACCCTGGAAGAGCCTGATGGAGGAGAGCTGGTCCTGGGGG
GCTCGGACCCGGCACACTACATNCCACCNCTCAACTTCGTGCCAGTCACGGTCCCGGNCT
ACTGGCAGATCCACATGGAGCGTGTGAAGGTGGGGNCCCAGGCTGACTTTNTNGTGGCCA
AGGGTGTGCTGNCCATC

SEQ ID NO 15

XFGXEGKLGSLGSHPAPQQCLHHCCCYPSLLLLPLLNVEPAGATLIRIPLRQVHPGRRTLNLLRGWGKPAELPKLGAPSP
GDKPASVPLSKFLDAQYFGEIGLGTPPQNFTVAFDTGSSNLWVPSRRCHFFSVPCWFHHRFNPNASSSFKPSGTKFAIQY
GTGRVDGILSEDKLTIGGIKGASVIFGEALWESSLVFTVSRPDGILGLGFPILSVEGVRPPLDVLVEQGLLDKPVFSFYF
NRDPEVVNGGELVLGGSDPAHYIPPLNFVPVTVPAYWQIHMERVKVGPRADSLCQGCAAILDTGTYLVITGPTEEIRALH
AAIGGIPLLAGEYIIRCSEIPKLPAVSLLIGGVWFNLTAQDYVIQTTRKGDVRLCLSGFRALDIAEAEGPVWILGEVFWG
ICDRFDRGT.RAANPS.LAGVALRGPXSGTEGXRGPQYR.RXSGLETYI

446 amino acids (+ 3 additional stop codons)

SEQ ID NO 16

NAATTCGGCANAGAAGGAAAACTAGGAAGCCTGGGTTCACACCCGGCTCCCCAGCAATGTCTCCACCACTGCTGCTGCTA
CCCTTCACTGCTGCTGCTGCCTCTCCTGAATGTGGAGCCTGCTGGGGCCACACTGATCCGGATCCCTCTTCGTCAAGTCC
ACCCTGGACGCAGGACCCTGAACCTACTGAGGGGATGGGGAAAACCAGCAGAGCTCCCCAAGTTGGGGGCCCCATCCCCT
GGGGACAAGCCTGCCTCGGTACCTCTCTCCAAATTCCTGGATGCCCAGTATTTTGGGGAAATTGGGCTGGGAACGCCTCC
ACAAAACTTCACTGTTGCCTTTGACACTGGCTCCTCCAATCTCTGGGTCCCGTCCAGGAGATGCCACTTCTTCAGTGTGC
CCTGCTGGTTCCACCACCGCTTCAATCCCAATGCCTCCAGCTCCTTCAAGCCCAGTGGGACCAAGTTTGCCATTCAGTAT
GGAACTGGGCGGGTAGATGGAATCCTGAGTGAGGACAAGCTGACTATTGGTGGAATCAAGGGTGCATCCGTGATTTTCGG
GGAAGCTCTGTGGGAATCCAGCCTGGTCTTCACTGTTTCCCGCCCCGATGGGATATTGGGCCTCGGTTTTCCCATTCTGT
CTGTGGAAGGAGTTCGGCCCCCGCTGGATGTACTGGTGGAGCAAGGGCTATTGGATAAGCCTGTCTTCTCCTTTTATTTC
AACAGGGACCCTGAAGTGGTTAATGGAGGAGAGCTGGTCCTGGGGGGCTCGGACCCGGCACACTACATCCCACCCCTCAA
CTTCGTGCCAGTCACGGTCCCCGCCTACTGGCAGATCCACATGGAGCGTGTGAAGGTGGGGCCCAGGGCTGACTCTCTGT
GCCAAGGGTGTGCTGCCATCCTGGATACAGGCACGTACCTGGTCATCACAGGACCCACTGAGGAGATCCGGGCCCTGCAT
GCAGCCATTGGGGGAATCCCCTTGCTGGCTGGGGAGTACATCATCCGGTGCTCAGAAATCCCAAAGCTCCCCGCAGTCTC
ACTCCTCATTGGGGGGGTCTGGTTTAATCTCACGGCCCAGGATTACGTCATCCAGACTACTCGAAAGGGTGACGTCCGCC
TCTGCTTGTCCGGCTTCAGGGCCTTGGACATCGCTCGGGCTGAAGGACCTGTCTGGATCCTCGGCGAAGTTTTTTGGGGA
ATATGTGACCGTTTTGACCGGGGGACATGAAGAGCGGCGAACCCGAGTTGACTTGCGGGGGTTGCCTTGCGAGGACCTTG
NAGCGGAACGGAAGGAGNCCGGGGTCCCCAGTATCGATGACGGNGAAGTGGTCTAGAAACATATATT

Claims

1. An isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity over its entire length to a nucleotide sequence encoding the polypeptide comprising the amino acid sequence of SEQ ID NO 1 or SEQ ID NO 15 or a fragment, analog or derivative of said polypeptide.

2. The polynucleotide of claim 1 which comprises the nucleotide sequence contained in SEQ ID NO 2 or SEQ ID NO 16.

3. An isolated polynucleotide which comprises a nucleotide sequence that is at least 80% identical to that of SEQ ID NO2 or SEQ ID NO 16 over its entire length.

4. The polynucleotide of claim 3 which is polynucleotide of SEQ ID NO: 1.

5. The polynucleotide of any one of claims 1 to 4 wherein the polynucleotide is DNA.

6. The polynucleotide of any one of claims 1 to 4 wherein the polynucleotide is RNA.

7. The polynucleotide of claim 5 wherein the polynucleotide is genomic DNA.

8. A vector containing the DNA of any one of claims 2, 4, 5, 6, or 7.

9. A host cell genetically engineered with the vector of claim 8.

10. A process for producing a polypeptide comprising expressing from the host cell of claim 9 the polypeptide encoded by said DNA.

11. A process for producing cells capable of expressing a polypeptide comprising genetically engineering cells with the vector of claim 8.

12. A polynucleotide hybridizable to the polynucleotide of any one of claims 1 to 7 and encoding a polypeptide having substantially the same biological function or activity as the polypeptide of SEQ ID NO 1 or SEQ ID NO 15.

13. A polypeptide which is at least 80% identical over its enire length to the amino acid sequence of SEQ ID NO 1 or SEQ ID NO 15 and fragments, analogs and derivatives thereof.

14. A polypeptide having the amino acid sequence of SEQ ID NO 1 or SEQ ID NO 15 and fragments, analogs and derivatives thereof.

15. The polypeptide of Claim 14 wherein the polypeptide has the amino acid sequence of SEQ ID NO 1 or SEQ ID NO 15.

16. The polypeptide of claim 13, 14 or 15 in isolated form.

17. The polypeptide of any one of claims 13 to 16 for use in therapy.

18. A method of screening compounds to identify those compounds which inhibit the polypeptide of claim 13, 14 or 15 comprising contacting isolated polypeptide with a test compound and measuring the rate of turnover of an enzyme substrate as compared with the rate of turnover in the absence of test compound.

19. A compound identified by the method of claim 18.

20. An inhibitor of the polypeptide of claim 13, 14 or 15.

21. An inhibitor according to claim 20 which is an antibody to the polypeptide of claim 13, 14 or 15.

22. A pharmaceutical composition comprising the polynucleotide of claim 1 or 12, a polypeptide of claim 13 or 14, a compound of claim 19 or an inhibitor of claim 20 and a pharmaceutically acceptable carrier.

23. A method for the treatment of a patient having need to inhibit the polypeptide of claim 13, 14 or 15 comprising: administering to the patient a therapeutically effective amount of the compound of claim 19 or inhibitor of claim 20.

24. The use of a compound of claim 19 or inhibitor of claim 20 for the manufacture of a medicament for use in therapy.

INTERNATIONAL SEARCH REPORT

Interna tional Application No
PCT/GB 97/02426

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/57 C12N9/64 C12N5/10 C12N15/11 A61K38/48
C12Q1/37 C07K16/40 A61K39/395 A61K31/70

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
A	P. FAUST ET AL: "Cloning and sequence analysis of cDNA for human cathepsin D" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 82, August 1985, WASHINGTON US, pages 4910-4914, XP002052494 ---	
A	N. BIRCH AND Y. LOH: "Cloning, sequencing and expression of rat cathepsin D" NUCLEIC ACIDS RESEARCH, vol. 18, no. 21, 1990, OXFORD GB, pages 6445-6446, XP002052495 --- -/--	

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
19 January 1998	20.02.98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Van der Schaal, C

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 97/02426

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	K. MORI ET AL: "Molecular cloning of a novel mouse aspartic protease-like protein that is expressed abundantly in the kidney" FEBS LETTERS., vol. 401, 20 January 1997, AMSTERDAM NL, pages 218-222, XP002052496 see the whole document ---	1-16, 19-21
P,A	TATNELL P J ET AL: "Cloning, expression and characterisation of murine procathepsin E." FEBS LETTERS 408 (1). 1997. 62-66. ISSN: 0014-5793, XP002052497 -----	

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 97/02426

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/GB 97/02426

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Remark : Although claim 23 is directed to a method of treatment of the human/animal body , the search has been carried out and based on the alleged effects of the compound/composition.